Plotting Maximum Likelihood for Right Censored Exponential Survival Times

Daniel Claborne

November 20, 2017

As the title suggests, I'm going to be looking at some right censored data and the likelihood for the distribution that produced it. The gist is that we want to see we have an unbiased estimator of the true parameter in the case of censored data.

Additionally, I will show that the formulas do not work in the presence of informative censoring.

Some handy functions

We create two functions that calculate the contribution to the overall likelihood for true and censored data. We scale the likelihood by the maximum value for plotting purposes. Since we are just dividing by a constant, this does not affect our maximum likelihood estimate.

```
###Returns a value of the likelihood for some value of the parameter (lambda), and data (x)
l_true <- function(lambda, x)


###Performs the same function but for censored data, with an added argument for censoring times
l_censored <- function(lambda, x, censor_time)
```

Additionally, we have a function that generates a dataframe of the true survival times and censored survival times

```
###Generates censored data of size n from some distribution (in our case, exponential), with the option
rsurv <- function(dist_function, n, censor_time = 1, parameter, censoring_mode = "fixed", ...)
```

Lets see what happens with fixed censoring...

We create two lists of 50 **samples** of size 50 from an exponential distribution with rate parameter $\lambda = 1$. The data is then censored at time 1.

```
censored_list <- list()
uncensored_list <- list()

for(i in 1:50){
  df <- rsurv(rexp, 50, censor_time = 1, parameter = 1)

  censored_list[[i]] <- df$censored_data
  uncensored_list[[i]] <- df$survival_times
}

rm(df)
```

Now we plot the likelihood functions for each sample.

```
####Code for first plot
mle1.plots <- map(uncensored_list, ~stat_function(fun = l_true, col = "green", args = list(.x)))
ggplot(data = data.frame(x = c(0,3)), mapping = aes(x = x)) +
            xlab("Lambda")+
```

1

```
ylab("Scaled Likelihood")+
ggtitle("Uncensored data, uncensored likelihood")+
mle1.plots
```

Uncensored data, uncensored likelihood

Censored data, censored likelihood

Censored data, uncensored likelihood
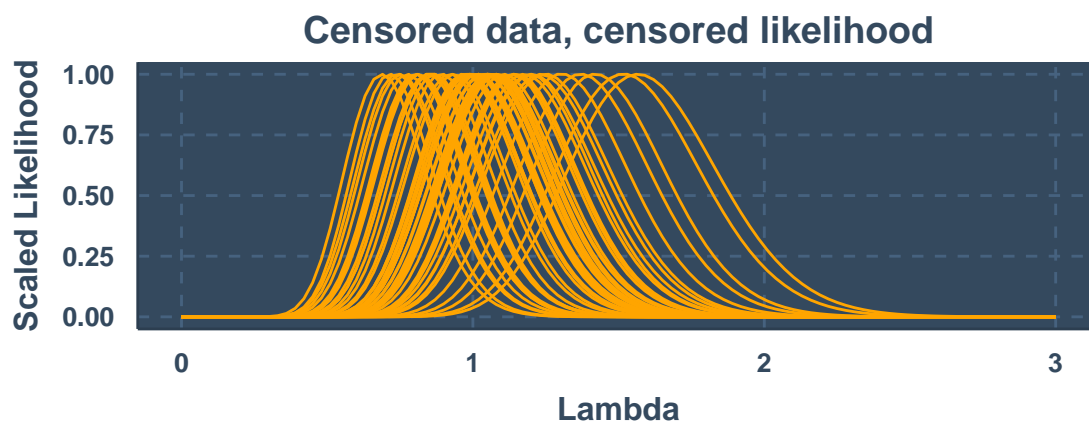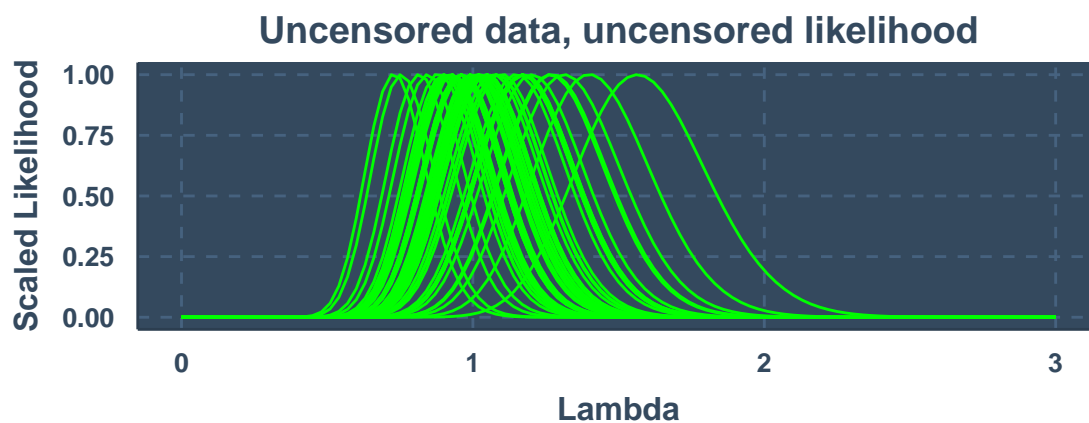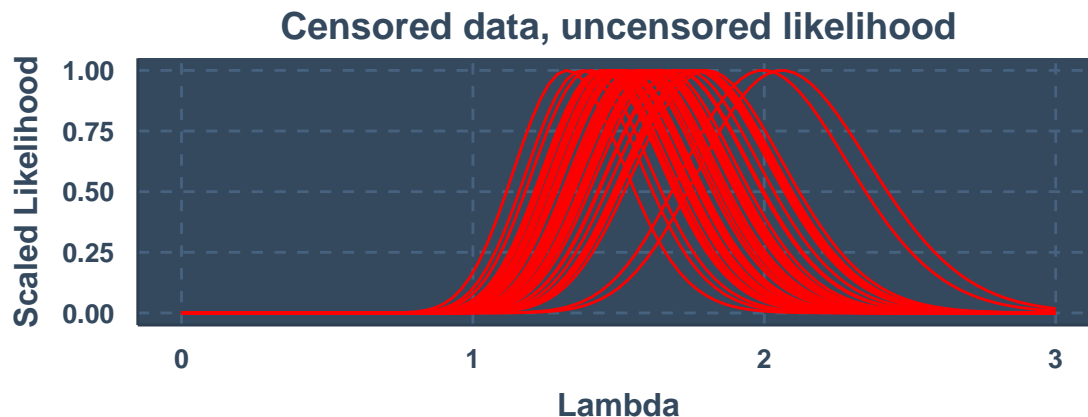
We can see that for uncensored and censored data, using the appropriate formula, the maximum of likelihood of λ is, across repeat samples, around the true value of the rate parameter ($\lambda = 1$). When we attempt to use the function for uncensored data on censored data, we can see that our estimates are biased towards larger values of λ. This makes sense, as with the rate paremetrization, the mean of an exponential distribution is $1/\lambda$. We have thus estimated a smaller mean survival time; a result of treating right censored(made smaller) observations as true event times.
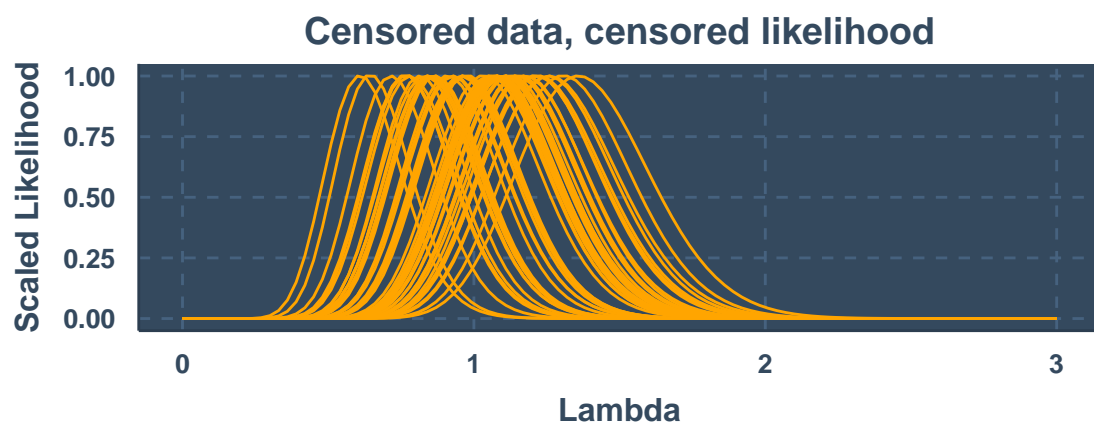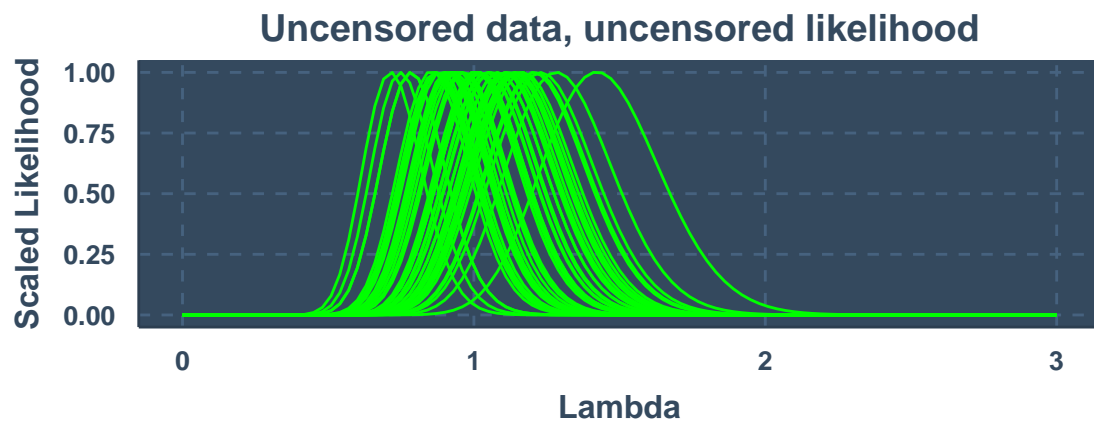
Random Censoring

If we employ random, noninformative censoring, we should still see unbiased estimates of λ, lets check:
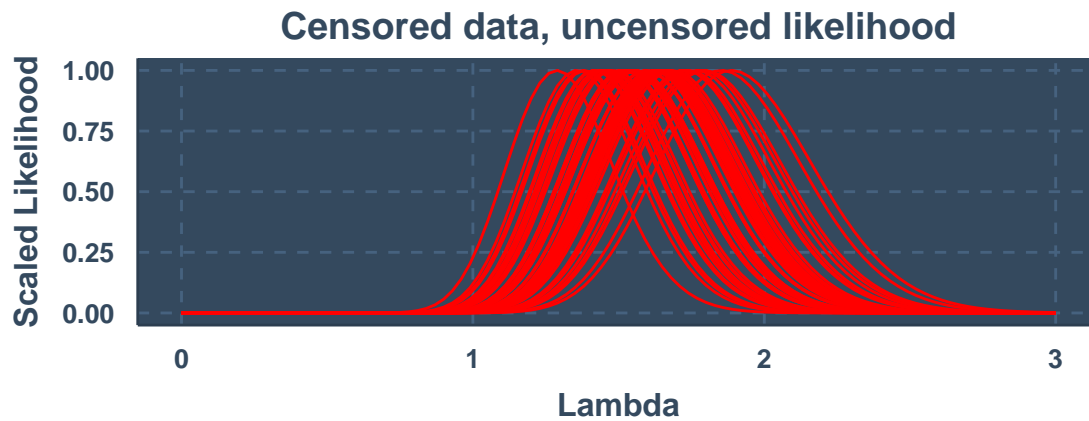
```r
censored_list <- list()
uncensored_list <- list()

for(i in 1:50){
  df <- rsurv(rexp, 50, parameter = 1, censoring_mode = "random")

  censored_list[[i]] <- df$censored_data
  uncensored_list[[i]] <- df$survival_times
}

rm(df)
```

Uncensored data, uncensored likelihood

Censored data, censored likelihood

Censored data, uncensored likelihood

We still see unbiased estimates for the true parameter as long as the appropriate likelihood is used.
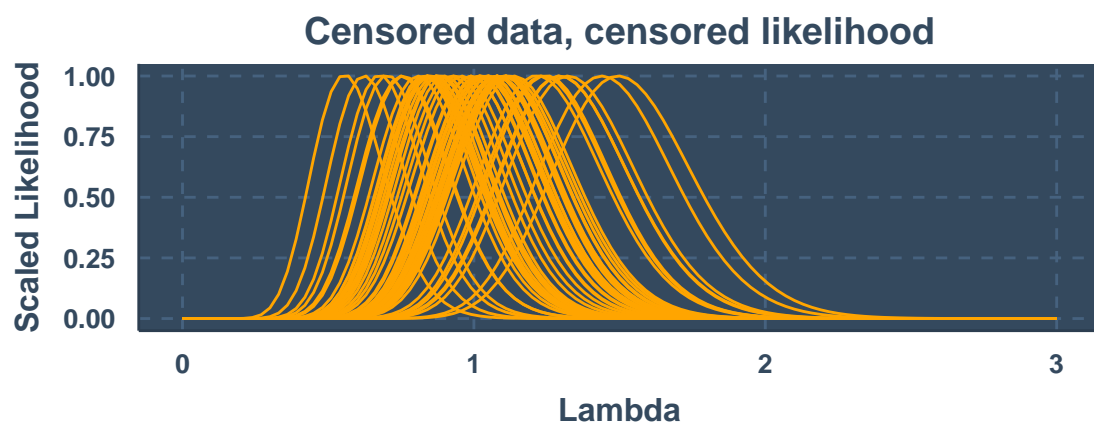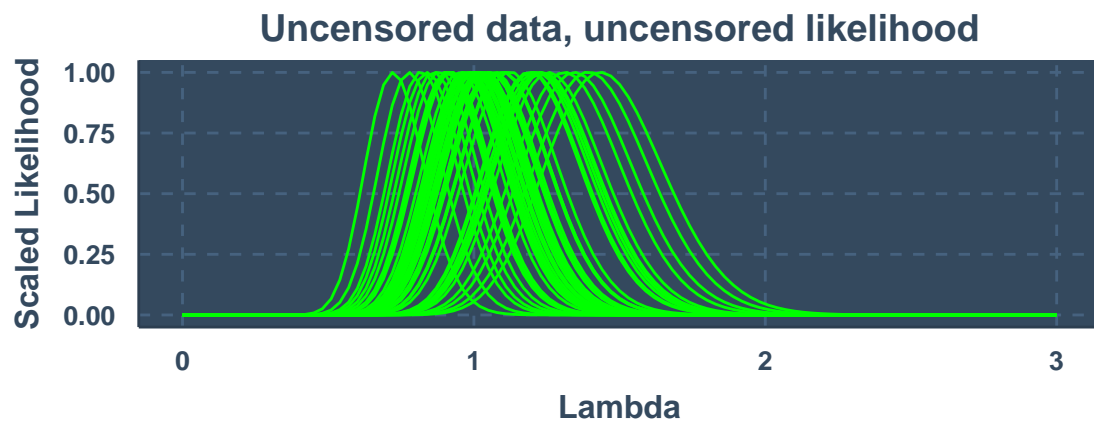
Informative Censoring

Very quickly, I will show that informative censoring invalidates the likelihood equations. The informative censoring happens by censoring survival times with different probabilities depending on whether or not the survival time is above or below the mean survival time.
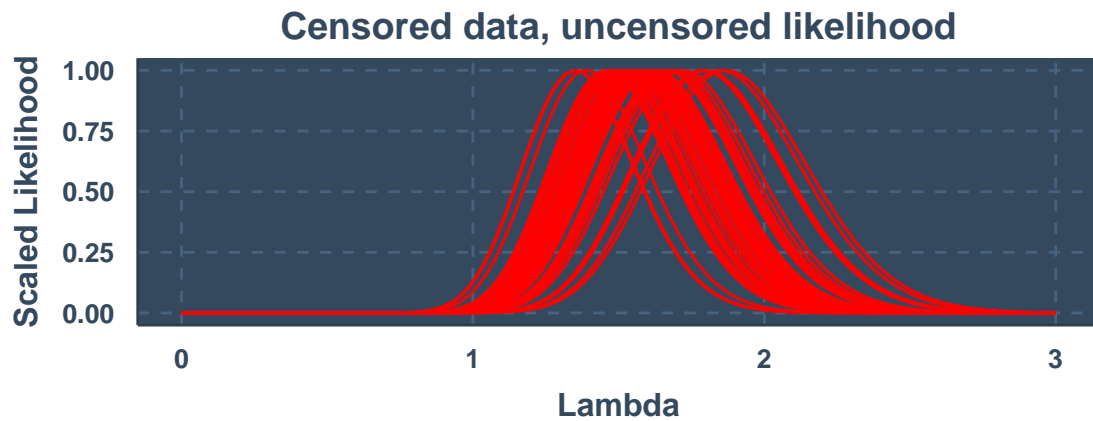
```r
censored_list <- list()
uncensored_list <- list()

for(i in 1:50){
  df <- rsurv(rexp, 50, parameter = 1, censoring_mode = "informative")

  censored_list[[i]] <- df$censored_data
  uncensored_list[[i]] <- df$survival_times
}

rm(df)
```

Uncensored data, uncensored likelihood

Censored data, censored likelihood

Censored data, uncensored likelihood

We see some positive bias in the maximum likelihood estimates, a result of the particular form of informative censoring I built into the function. **Note:** All three graphs show a biased estimate, you can eyeball a rightward shift even in the green and orange plots as compared to the same plots in the fixed and random censoring cases.

In summary, when no censoring occurs, the likelihood is easily computed we can see that the MLE is an unbiased estimator of the true parameter. When there is censoring, either fixed or truly random, we must make sure that contributions to the likelihood from censored observations account for (use the second formula/function) the censoring. In the case of right censored data, failure to do so will result in an upward bias in our estimates. Finally, we saw that in the presence of informative censoring, the MLE is biased even when we account for the censoring in the likelihood.